Computational Tools for Reconstruction and Discovery of Metabolic, Signal transduction, And Evolutionary Pathways

Andrew Pohorille, NASA-ARC

Biomolecular Systems Research Program

Our high-level research objective is to develop computational tools that aid biologists in forming and evaluating explanations of cellular processes in a semi-automated manner. These tools will be particularly suitable for reconstruction and discovery of metabolic, signaling, and evolutionary pathways. They will be necessary if biologists are to take full advantage of the newly acquired ability to collect large amounts of data through genomic and microarray studies. Further, we will demonstrate capabilities of this software in studies on problems of current biological interest, such as the nature of metabolic processes, gene regulation, and the evolution of unicellular organisms. The proposed work will contribute to improving our understanding of living systems, their evolution, and their adaptability both on Earth and elsewhere.

Description

1. A formalism and programming language for expressing models of biological processes at varying levels of abstraction, capable of supporting both qualitative and quantitative descriptions.
2. Simulation software that operates on this formalism to make qualitative and quantitative predictions about the behavior of a biological system under specific environmental conditions.
3. Explanatory software that generates candidate reactions and reaction pathways to account for detected anomalies or resolve encountered difficulties arising from missing, inaccurate or conflicting data, with the explanations constrained by knowledge of biochemical plausibility.
4. Evaluation software that ranks candidate reactions and pathways in terms of their ability to fit data and their consistency with background knowledge of biochemical, DNA sequence, and protein data.
5. Theory refinement software that proposes revisions to the current model based on these explanations and, subject to approval by the biologist, adds them to the knowledge base.

Innovative Claims/NASA Significance

Although our approach builds on earlier work in bioinformatics, knowledge representation, and computational scientific discovery, it makes some novel and important contributions:

1. The formalism for expressing biological processes supports models at both the traditional concrete levels (e.g., specific biochemical reactions) and more abstract levels (e.g., signal transduction, energy generation, and evolutionary processes).
2. The software moves beyond simulation and prediction to compare predictions with data and, more important, generates candidate explanations that are consistent with existing biological knowledge.
3. Both the formalism and the software will be natural for biologists to use, and they will let biologists understand their data in terms already familiar to them and extend their knowledge in ways that they find acceptable.

In general, the software we develop will let biologists deal with the increasing amounts of bioinformatic data while taking advantage of their considerable and constantly improving knowledge about biological processes.

Plans

1. After year 1 - release the alpha-version of Biolingua with capabilities for qualitative and quantitative metabolism reconstruction and for bridging missing reactions.
2. After year 2. - (a) complete validation of the metabolism part of Biolingua using biochemical and microarray data, and (b) release the alpha-version of Biolingua capable of supporting abstract processes needed for reconstructing signal transduction pathways and gene regulation networks.
3. After year 3 - release web-based Biolingua software with full capabilities to study metabolic, regulatory and evolutionary processes.